FORMING OF LEARNING SET FOR NEURAL NETWORKS IN PROBLEMS OF LOSLESS DATA COMPRESSION

Yuriy Ivaskiv, Victor Levchenko

Abstract: questions of forming of learning sets for artificial neural networks in problems of losless data compression are considered. Methods of construction and use of learning sets are studied. The way of forming of learning set during training an artificial neural network on the data stream is offered.

Keywords: neural network; modelling of data sources; learning set.

ACM Classification Keywords: C.1.3 Other Architecture Styles – Neural Nets. D.4.8 Performance – Modelling and predictions. E.4 Coding and Information Theory – Data compaction and compression.

Introduction

In problems of data compression wide application was found with artificial neural networks (ANN). The neural methods of losless data compression have got wide spread occurrence [Jiang, 1999]. Applicability of lossy compression methods is limited to a class of multimedia data, and neural realization of such methods is specific to each concrete type of data that limits an opportunity of their use [Cottrell, 1987; Verma, 1999]. Methods of losless data compression differ that at their use, as a rule, special restrictions on structure and type of data are not imposed that expands opportunities of their practical application for the decision of various classes of problems.

The most of modern methods of losless data compression are entropic, that is such in which the length of a code necessary for representation of an element of data, contacts probability of occurrence of an element in a stream of compressed data [Ватолин, 2002]. Division of a problem of compression into two problems is standard: *coding* and *statistical modelling* of a data source [Rissanen, 1981]. Now various methods and algorithms of coding redundancy of which practically does not differ from theoretically minimal possible are developed and applied [Кричевский, 1989]. However, a problem of increase of efficiency of statistical modelling is insufficiently studied and an opportunity of increase of compression ratio in particular is connected with its decision.

The existing statistical models used for data compression, differ complexity and heterogeneity of structure, and also that they are sensitive to type of modelled data that complicates their practical realization and application [Bell, 1989; Witten, 1989; Ker-Chang Chang, 1993]. At the same time the models which are based on ANN, differ simplicity and uniformity of structure, an opportunity of use of the same algorithms of adjustment and functioning in various models, and also some other useful properties [Хайкин, 2006]. Therefore, prospects of development of statistical models are connected with use as modelling system at losless compression of the artificial neural networks.

Target of Settings

The problem of statistical modelling of data sources during realization of losless compression consists in definition of probabilities of occurrence of elements of compressed data on an input of coding system. One of methods of reception of estimations of these probabilities is forecasting occurrence of elements of data in an input stream. At use ANN such forecasting is carried out similarly to forecasting of time series [Ежов, 1998].

Distinction consists basically in principle of forming of learning set for ANN. At forecasting time series there are accessible all values that enables training of ANN with the given level of an error. At statistical modelling data come in an input of modelling system in the consecutive portions that demands permanent recustomizing of model during its functioning with the purpose of adaptation to changes of data [Schmidhuber, 1996]. In this connection there is a necessity of development and research of special methods and algorithms of formation of learning set for neural models which would consider features of its functioning on the given data stream.

Main Results

Description of neural model.

For carrying out of researches the two-layer back propagation neural network with adjustable quantity of neurons in an input layer (if the inputs of ANN are considered as the first layer such ANN is considered to be three-layer) was used. As one of possible variants of realization of transferring function the hyperbolic tangent – *th x* – was used. Other parameters ANN had following values:

- the maximal absolute error of training per one example: 0,001;
- a steepness of transfer function: 1,0 – for an input layer, 0,1 – for a target layer;
- factor of learning speed: 0,1 – for all layers.

Proceeding from convenience of realization neural models on the existing computer-based means dimension of 1 byte has been accepted for elements of the input data stream that is not basic restriction of parameters of considered model. Accordingly, the quantity ANN inputs was defined as 8N, where N – quantity of data elements on which forecasting (that is context length), and the quantity of outputs was fixed and was equal 8 (fig. 1). Coding of inputs and target signals of a network has great value for process of ANN learning because finally it defines finish of learning. Zero bits of inputs and target signals of a network were coded by value "-0,5", and set bits – "+0,5" as use of such values allows to consider process of learning completed only at concurrence of signs of the received and required values of a network without achievement of a concrete numerical result.

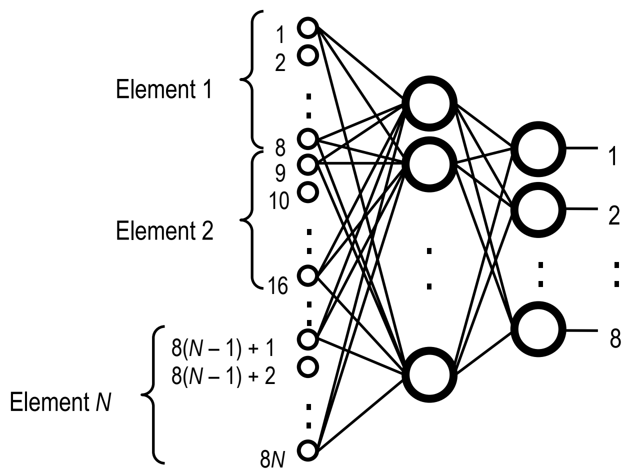

Fig. 1

Influence of context length.

During ANN functioning it was learned of forecasting of the values appearing in an input data stream on contexts with length from 1 up to 15 elements. Here start value answers minimally possible length of a context and the range of its change is defined by known methods of contextual modelling [Ватолин, 2002]. It is established, that quality of forecasting in a greater measure depends on length of a context for the data, described by strongly pronounced internal structure (texts in natural languages and texts of programs) and in a smaller measure – for data of type "an analog signal" (the digitized images and a sound). For text data dependence of optimum length of a context on volume of data is revealed that is coordinated with experimental data. At use of contexts of optimum length the quantity of correct forecasts has increased on the average on (25 – 45) %.

Window learning mode.

As a possible way of increase of quantity of correct forecasts window mode of ANN learning is considered at which learning is carried out not on one context, but on set of the overlapped consecutive contexts that are directly previous of prediction value (fig. 2). Researches were spent for windows with a width from 2 up to 50 contexts. The range of change width of a window is chosen experimentally. The basic problem of formation of learning set for a window mode of learning was elimination of inconsistent contexts [Тарасенко, 2001], the presence of which cycles process of learning and decreases speed of functioning of model sharply. Elimination of inconsistent contexts was carried out as follows: if there were two or more conterminous contexts with various prediction values in a window the nearest to prediction value was left and all previous were removed.

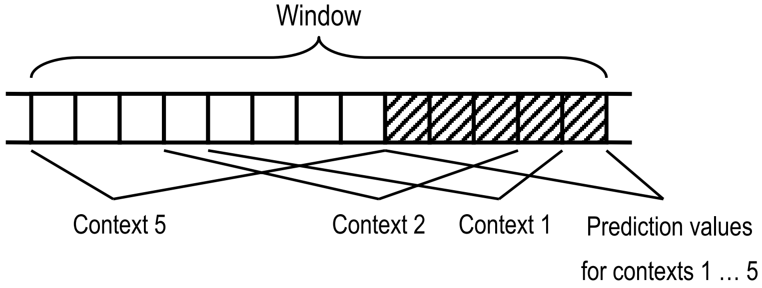

Fig. 2

As a result of test of neural model in a window mode of learning it has been established, that the increase in width of a window allows raising quantity of correct forecasts (on 40 % for text data), however it reduces speed of ANN learning, and speed of model as a whole. Besides the increase in length of a context leads to reduction of number of the correct forecasts connected with increase of width of a window (fig. 3). Thus the window mode of learning is favourable only at use of contexts of small length (1 … 4 elements) under condition of elimination of inconsistent contexts.

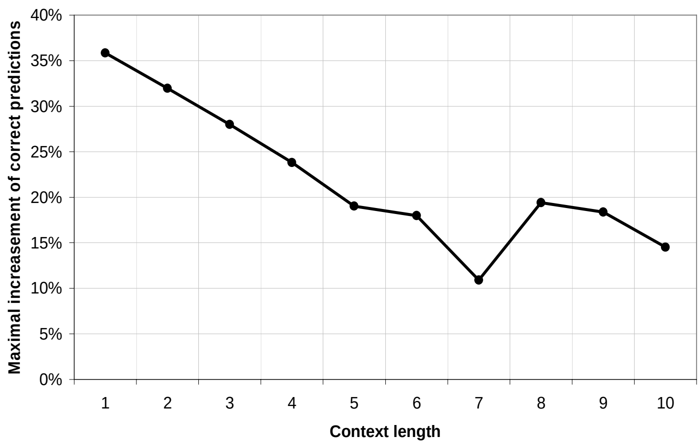

Fig. 3

Selective learning

The analysis of change of quality of forecasting during model functioning has shown, that neural model functioning on all set of contexts with any width of a window is characterized by initial splash in quantity of correct forecasts and its following decrease to some level. It is supposed, that such behaviour of quality of forecasting shows that only some subset from all contexts used for ANN learning represents laws, characteristic for all set of data, and the others contexts have casual character and lead to

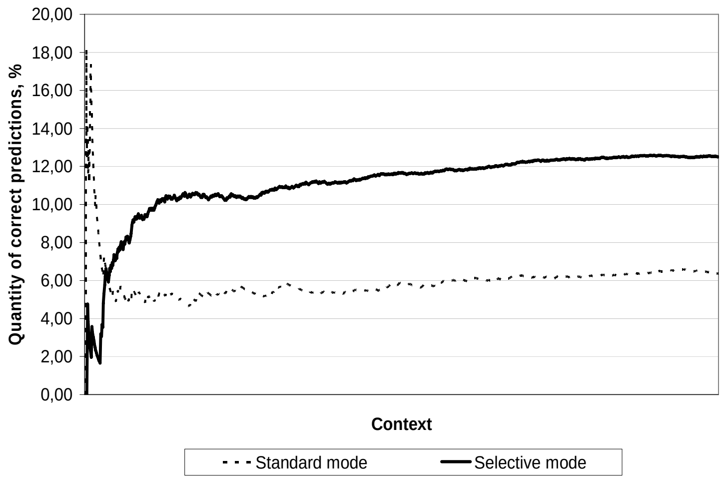

Fig. 4

decrease in quality of training process. For check of the given assumption learning of a network has been led in two modes: 1) standard (on all set of contexts); 2) selective (on set of contexts after which forecasting was successful). As a result it was revealed, that at selective learning change of quantity of correct forecasts was absolutely different unlike a standard mode (fig. 4); thus the stable increase in quality of forecasting during all process of ANN training and functioning is observed.

Essential lack of a selective learning mode is dependence of quality of forecasting process on an initial condition of model, which is on a choice of initial values of ANN weight factors. The reason of such dependence is that at a selective mode training of a network begins with the first context for which the correct forecast has been executed, thus the probability of such forecast depends on initial values of ANN weight factors. In connection with that casual values are usually appropriated to ANN weight factors, there is a probability of their unsuccessful choice, as entails sharp decrease in speed, and quality of ANN learning also.

The account of influence of an initial condition of model on forecasting process is offered to be carried out by one of the next two ways. The first way assumes to carry out learning of a network on contexts for which the forecast has been executed with the error that is not exceeding given. Thus, in view of that modelled data are represented by binary codes, for an estimation of an error it is expedient to use Hemming distance. At a choice of Hemming distance it is important to consider, that when admissible error of forecasting increases, the quantity of the casual contexts increases also and it makes worse quality of forecasting, and when admissible error of forecasting decreases, the probability of the duly beginning of ANN learning decreases also.

The second way assumes application of the mixed mode of learning at which in the beginning of ANN functioning it is trained in a usual mode, which is on all meeting contexts (“start” of model), and after a time learning becomes selective. The choice of length of data on which “start” of model is carried out depends on type of data, and also their structure, defined by a solved problem and demands special studying.

Conclusion

The choice of a context length for learning set should be carried out in view of type of modelled data.

The choice of a context length without taking into account type of modelled data does not allow getting optimum quantity of correct forecasts that can be compensated by increase in width of a learning window.

Forming of training set from small contexts (1 … 4 elements), and also with use of windows in width from 7 up to 10 contexts is preferable to increase in number of correct forecasts at the solving of problems of losless compression for various types of data.

The greatest number of correct forecasts in the data stream can be achieved at the use of a selective mode of learning with preliminary “start” of model or with use of forecasts within the limits of Hemming distance no more than 2 bits.

Bibliography

[Jiang, 1999] Jiang J. Image compression with neural networks – A survey. Signal Processing: Image Commun. 1999. – V. 14. – P. 737 – 760.

[Cottrell, 1987] Cottrell G. V., Munro P., Zipser D. Image compression by back propagation: An example of extentional programming // Proc. 9th Annual Confer., Cognitive Soc. – 1987. – P. 461 – 473.

[Verma, 1999] B. Verma, M. Blumenstein, S. Kulkarni. A New Compression Technique Using an Artifical Neural Network // Journal of Intelligent Systems. – 1999. – Vol. 9. – P. 39 – 53.

[Ватолин, 2002] Ватолин Д., Ратушняк А., Смирнов М., Юкин В. Методы сжатия данных. Устройство архиваторов, сжатие изображений и видео. – М.: ДИАЛОГ-МИФИ, 2002. – 384 с.

[Rissanen, 1981] Jorma Rissanen, Glen G. Langdon. Universal modelling and coding. // IEEE transaction on information theory. 1981. V. IT – 27. № 1. – P. 12 – 33.

[Кричевский, 1989] Кричевский Р. Е. Сжатие и поиск информации. – М.: Радио и связь, 1989. – 168 с.

[Bell, 1989] Bell T. C., Moffat A. M. A note on the DMC data compression scheme. // Comput. J. – 1989. – V. 32. – P.16–20.

[Witten, 1989] Bell T. C., Witten I. H., Cleary J. G. Modeling for text compression // ACM Computer Surveys. – 1989. V. 24. – P. 555 – 591.

[Ker-Chang Chang, 1993] H. Ker-Chang Chang, Shing-Hong Chen. A New Locally Adaptive Data Compression Scheme Using Multilist Structure // Comput. J. – 1993. V. 36.

[Хайкин, 2006] Хайкин С. Нейронные сети: полный курс, 2-е издание.: Пер. с англ. – М.: Издательский дом “Вильямс”, 2006. – 1104 с.

[Ежов, 1998] Ежов А. А., Шумский С. А. Нейрокомпьютинг и его применение в экономике и бизнесе (серия “Учебники экономико-аналитического института МИФИ” под ред. проф. В. В. Харитонова). – М.: МИФИ, 1998. – 244 с.

[Schmidhuber, 1996] Schmidhuber J. Sequential neural text compression // IEEE transaction on Neural Networks. – 1996. – V. 7 (1). – P. 142 – 146.

[Тарасенко, 2001] Тарасенко Р. А., Крисилов В. А. Предварительная оценка качества обучающей выборки для нейронных сетей в задачах прогнозирования временных рядов. // Тр. Одес. политехн. ун-та. – Одесса, 2001. – Вып. 1. С. 25 – 28.

Authors' Information

Yu. L. Ivaskiv - National Aviation University, Doctor of Science; professor; P.O.Box: 03058, Kosmonavta Komarova Ave., bl. 1, Kiev, Ukraine; e-mail: post@nau.edu.ua.

V. V. Levchenko - Industrially-Economic College of National Aviation University, post-graduate student; P.O.Box: 03126, Geroev Sevastopolya St., bl. 44 - A, 53, Kiev, Ukraine; e-mail: vvl@voliacable.com